7th Annual Epidemiology, Biostatistics and Clinical Research Methods Summer Session

June 20-24, 2005

Using VA Databases For Research: Focus on Cancer

Denise M. Hynes, PhD, RN

Min-Woong Sohn, PhD

VIReC RESEARCHERS' GUIDE TO VA DATA

Welcome

- Course faculty
- Learning objectives
- Course outline
- Recommended readings
- Bibliography
- Housekeeping details

Learning Objectives

- Know about different types of VA data & information systems
- Become aware of potential uses of VA data for research
- Understand limitations of VA information systems data for research, especially in cancer
- Become aware of past and potential research applications in cancer
- Know about resources to support research use of VA data

This Course covers…

- Day 1: Introduction to VA Data & Research Uses
  - Overview VA Information Systems & National Databases
  - Good Data Practices: Privacy & Security
  - VA Inpatient & Outpatient Care Datasets
  - Prostate Cancer Example

- Day 2: Overview of VA Clinical Databases
  - VA DSS Clinical National Data Extracts
  - VA Pharmacy Data
  - Research Examples

This Course covers…

- Day 3: Using VA and non-VA Data
  - Medicare Data
  - Cancer Registry Data
  - VA & non-VA Mortality Data Sources

- Day 4: Focus on Cancer Research Applications & Programming Examples
  - Colon Cancer Treatment in VA & Medicare
  - Accessing the VA AAC Databases with Programming Examples

7th Annual Epidemiology, Biostatistics and Clinical Research Methods Summer Session

June 20-24, 2005

Introduction to VA Data

VIReC
RESEARCHERS' GUIDE TO VA DATA

Session Objectives

- Become aware of utility of VA data sources for cancer epidemiology and health services research
- Become aware of the scope and breadth of VA data
- Know about *G*ood *D*ata *P*ractices
- Understand how to use the VA Inpatient & Outpatient datasets for research
- Describe an example of determining cancer prevalence using VA data
- Know where to go for help when using VA data

Session Objectives

- **Become aware of utility of VA data sources for cancer epidemiology and health services research**
- **Become aware of the scope and breadth of VA data**
- **Know about *G*ood *D*ata *P*ractices**
- **Understand how to use the VA Inpatient & Outpatient datasets for research**
- **Describe an example of determining cancer prevalence using VA data**
- **Know where to go for help when using VA data**

Relevance of VA Data to Cancer Research

- Identifying prevalence and incidence of specific cancers in the VA population
- Rates of screening for any type of cancer
- Stages at diagnosis for any type of cancer
- Rates of treatment vs. non-treatment, where cancer treatment is known to be effective
- Disparities in cancer screening or treatment
- Recurrence rates, survival rates, complication rates, and other outcome measures
- Disease management in cancer care
- Cancer care treatment patterns across systems of care
- Clinical trials in cancer treatment involving VA patients
- Implementation of best practices in cancer care

Research Examples:
Wilt et al., *Med Care* 1999

- Utilization and mortality outcomes of radical prostatectomy (RP)
- Used VA Inpatient & Outpatient Medical SAS Datasets from 1986 thru 1996 to identify
  - Patient cohort with RP
  - Complications
  - Comorbid conditions
- Used VA BIRLS Death File to obtain vital status
- Utilization of RP more than doubled between 1986 and 1996
- Tremendous geographic variations
  - Utilization rates lower in Eastern states
  - Variations decreased over time
- 30-day mortality decreased over time

Research Examples:
Fisher et al., *AJG* 2003

- Effect of follow-up colonoscopy on mortality
- New colorectal cancer cases during 1995 – 1996
- Used VA Medical SAS Inpatient Datasets for:
  - Patient identification
  - Comorbidity status using Deyo-Charlson method
- Used VA Medical SAS Inpatient and Outpatient Datasets
  - Colonoscopies
  - Outpatient visits
  - Chemotherapy or radiation therapy
- BIRLS Death File to obtain vital status
- Risk of death decreased by 43% with follow-up colonoscopy

Research Examples:
Rabeneck et al., *AJG* 2004

- Hospital surgical volume and long-term survival
- Used VA Medical SAS Inpatient Datasets
  - Patient cohort – surgical resection in 1991 – 2000
  - Comorbidity status using Deyo-Charlson method
  - Surgical volume
- Used VA Medical SAS Inpatient and Outpatient Datasets
  - Demographic data
  - Chemotherapy & radiation therapy
- BIRLS Death File to obtain vital status
- High surgical volume assoc. w/increased 5-yr survival
  - 7% for colon cancer
  - 11% for rectal cancer

Session Objectives

- Become aware of utility of VA data sources for cancer epidemiology and health services research
- **Become aware of the scope and breadth of VA data**
- Know about *G*ood *D*ata *P*ractices
- Understand how to use the VA Inpatient & Outpatient datasets for research
- Describe an example of determining cancer prevalence using VA data
- Know where to go for help when using VA data

Department of Veterans Affairs (VA)

- Provides federal benefits to veterans and dependents
  - Health Care (Veterans Health Administration)
  - Benefits (Veterans Benefits Administration)
  - Cemeteries (National Cemetery System)

- FY2005 estimated budget = $67 billion

Veterans Health Administration (VHA)

In 2005:

- 21 Networks
- 158 hospitals
- 132 nursing homes
- 42 domiciliaries
- 854 outpatient clinics
- 88 comprehensive home-care programs

Source: Facts About the Department of Veterans Affairs, at http://www1.va.gov/opa/fact/vafacts.html, accessed April 1, 2005.

VA Research & Development Service

- Biomedical Laboratory R&D Service
- Clinical Science R&D Service
- Health Services R&D Service
- Rehabilitation R&D Service

Who Can Use VA Data For Research?

- Employment status
- Purpose
- Ownership/Management/Authorizations
- Physical Location/Format
- Sensitivity of the information

Sources of VA Data Available

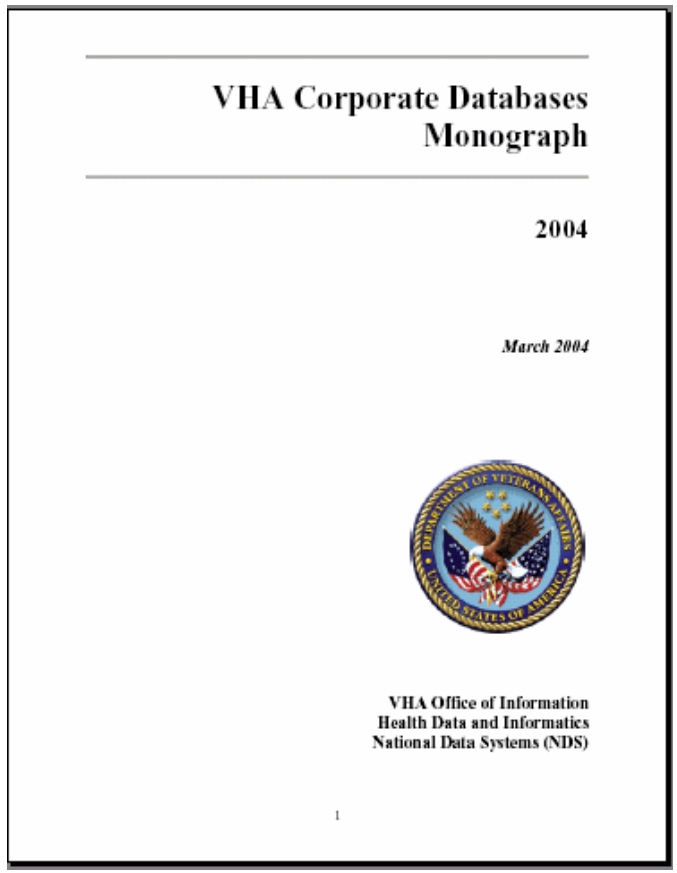
- Administrative/operations data
- Electronic medical record information
- Patient-derived data

Levels of Data

- Local Facility Level
  - Information may reside only at local facility
- Corporate (National) Level
  - Mandate for some local data may include uploading a standardized component to a central location
- VA Network Level: VISN Warehouses
  - Above local level, below corporate level

Corporate Databases Monograph

- Produced by OI National Data Systems
- http://www.virec.research.med.va.gov



VHA Corporate Databases Monograph

2004

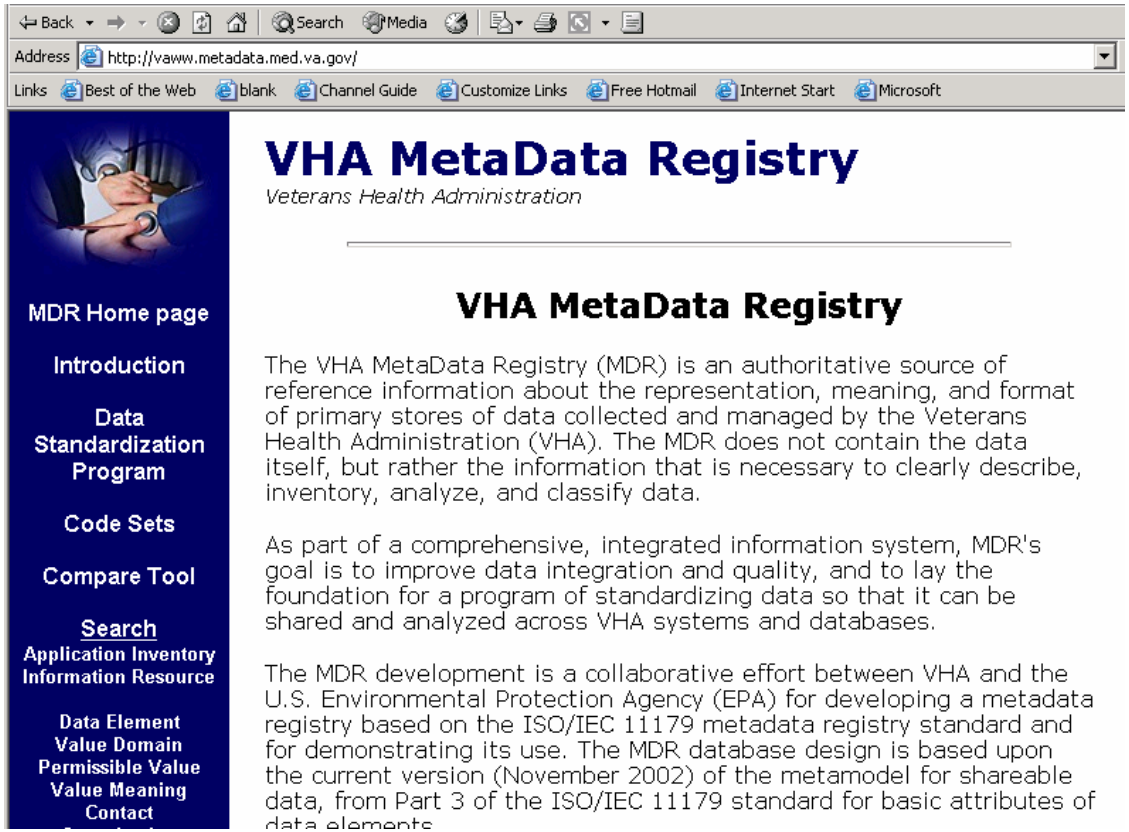
March 2004

VHA Office of Information
Health Data and Informatics
National Data Systems (NDS)

1

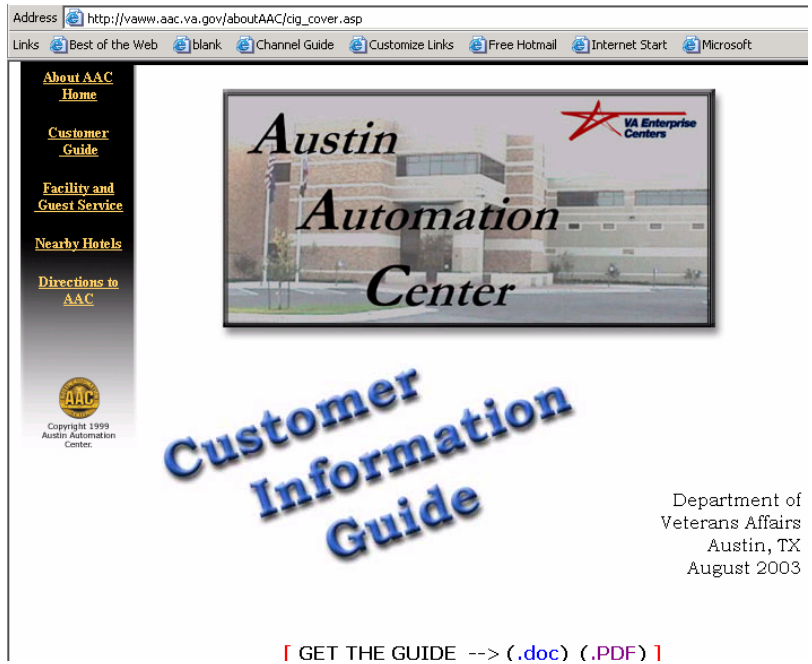
Search Meta-Data Registry

- http://vaww.metadata.med.va.gov



AAC Customer Information Guide

- http://vaww.aac.va.gov/aboutAAC/cig_cover.asp
- Applications, tools, utilities, & online facilities
- Contact information for stewards & programmers



Separate Presentations During The Course

- VA Inpatient & Outpatient Data
- VA DSS Clinical National Data Extracts
- VA Pharmacy Data
- VA Linked Medicare Data
- VA & non-VA Mortality Data
- VA & NCI Tumor Registry Data

Brief Remarks on Some Other VA Databases of Interest

- Geographic Data
- Economic Data from HERC
- VA Enrollment Data
- Survey of Veterans
- Large Health Survey
- OQP Survey of Health Experiences of Patients
- Health Data Repository

Planning Systems Support Group Geographic & Facility Data

- VHA Office within ADUSH
- Maintains VA Site Tracking (VAST) Data
  - Some organizational data on all VHA facilities
  - No yearly files
  - Updated info with change histories
  - Crosswalk file useful to track changes in facility IDs
- Geographic Information System (GIS) data on VA facilities, time-to-travel data, and ZIP Code centroid data

HERC Average Cost Datasets

- Acute hospitalizations, non-acute hospitalizations, long term care stays, outpatient care
- Housed at AAC
- Contact HERC to request ACRS Functional Task Code

- See http://www.herc.research.med.va.gov

Enrollment Database

- Health Eligibility Center (HEC)
  - Form 10-10EZ, Application For Health Benefits: https://www.1010ez.med.va.gov/sec/vha/1010ez/Form/vha-10-10ez.pdf
  - Scrambled SSN, Sex, Age, Marital Status,Veterans' Addresses, Phone Numbers, SC Percent, SC (Yes/No), POW
  - Race/ethnicity – new items in Form 10-10EZ
  - Eligibility Status (*e.g.*, cancelled, declined, deceased, not eligible) and Priority.
- NED (National Enrollment Database) will be replaced by EDB (Enrollment Database)
- SAS datasets at AAC
  - Extract of NED

Survey of Veterans

- VA Office of Policy, Planning, & Preparedness
- Periodic national phone surveys of all veterans
  - 1979, 1983, 1987, 1992, 2001
- Demographics, socio-economic status
- Military service experience, VA benefits
- Health status, insurance, utilization
- 2001 methods and results at
  http://www.va.gov/vetdata/SurveyResults/index.htm
- SAS files at AAC with access via ACRS

Survey of the Health Experiences of Patients (SHEP)

- Office of Quality & Performance
- Mail surveys
- Ambulatory Care & Inpatient Care SHEP (monthly)
- Veteran Satisfaction Survey for Prosthetics/ Sensory Aids (every other year)
- See details & data request procedures at http://vaww.oqp.med.va.gov/

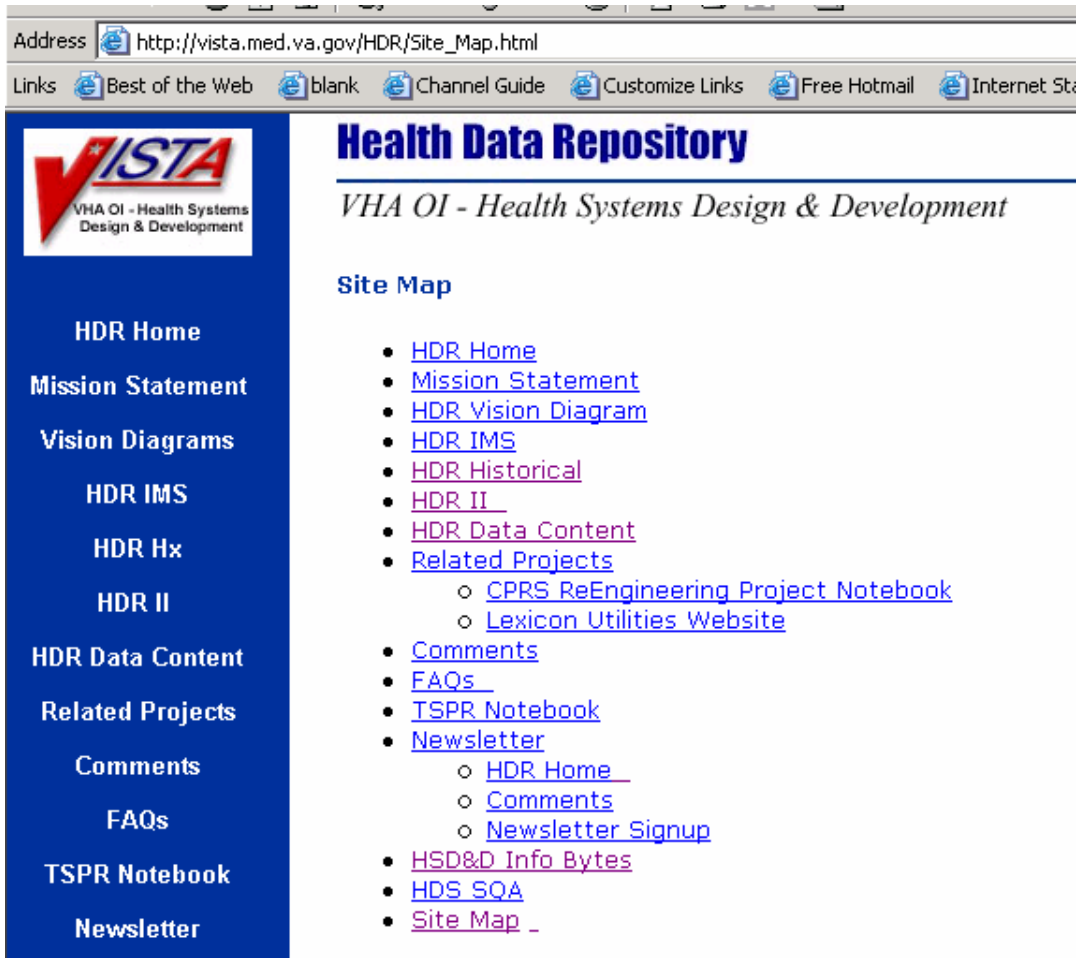
Financial & Clinical Data Mart (FCDM)

- Integrated patient, clinical, & financial data at facility & VISN level
- VISN Support Services Center (VSSC), Assistant Deputy Under Secretary for Health (ADUSH)
- Accessible via Intranet
- ProClarity Software
  - Data cubes→Graphs, reports, drill-down
- See http://klfmenu.med.va.gov

Health Data Repository

- National clinical data warehouse
- VHA OI Health Systems Design & Development
- Purposes:
  1. Primary source for the legal medical record
  2. Reports based on the entire clinical holdings of VHA
  3. Platform for a re-engineered CPRS
  4. Platform for patient self-access to medical record
  5. Part of standardization between & among Department of Defense, Indian Health Services, other government and private industry clinical databases
- VistA data 1999 & after by July 2005

See The Future Unfold at http://vista.med.va.gov/HDR



Session Objectives

- Become aware of utility of VA data sources for cancer epidemiology and health services research
- Become aware of the scope and breadth of VA data
- **Know about *G*ood *D*ata *P*ractices**
- Understand how to use the VA Inpatient & Outpatient datasets for research
- Describe an example of determining cancer prevalence using VA data
- Know where to go for help when using VA data

Privacy & Security Issues

- VA Upholds Patients' Rights
- Policy and procedures ensure ADP Security
- Privacy Act and Data Security Statement
- Medical Information Security Service (MISS)
  - (193C)
  - Phone: (304) 262-7300
- VHA Directive 6210, March 7, 2000
  - "Automated Information Systems (AIS) Security"

Research Use of VA Data

- Use or disclosure of VHA patient data is governed by a series of laws and regulations—not just HIPAA regulations
- Researchers' use generally must have IRB approval—with a few exceptions such as work preparatory to research
- Research requests for real SSN go through VACO CRADO
  - Real SSN is just one of many PHI to protect

What Is HIPAA?

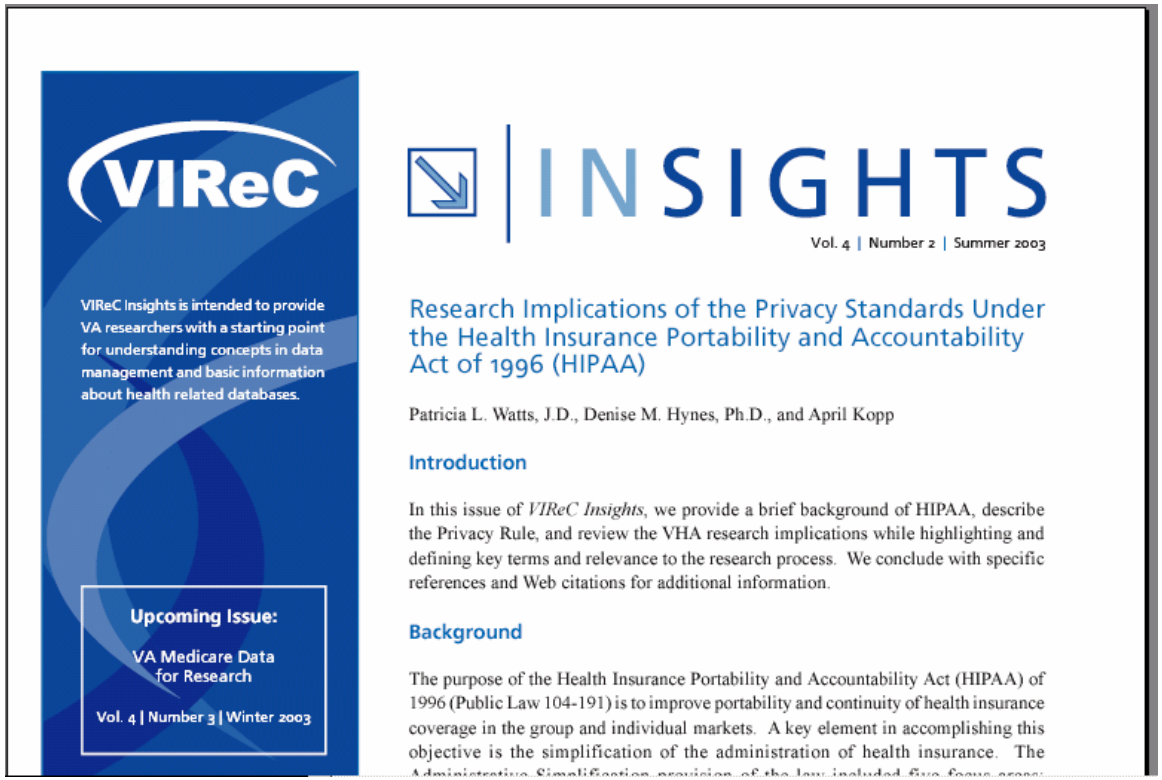
- Health Insurance Portability & Accountability Act of 1996 (PL 104-191)
  - Enacted August 21, 1996

- Intent: To assure health insurance portability, reduce healthcare fraud and abuse, guarantee security and privacy of health information, and enforce standards for health information

HIPAA – cont'd

- Privacy Rule
  - Compliance date April 14, 2003
  - Implementation ongoing
- Security Rule
  - Effective date: April 21, **2003**
  - Compliance date: April 21, **2005**
- VHA Activities: HIPAA PMO/OCIS
  - Gap analysis; implementation for Security Rule

Research Implications of HIPAA



VIReC

VIReC Insights is intended to provide VA researchers with a starting point for understanding concepts in data management and basic information about health related databases.

Upcoming Issue:

VA Medicare Data for Research

Vol. 4 | Number 3 | Winter 2003

INSIGHTS

Vol. 4 | Number 2 | Summer 2003

Research Implications of the Privacy Standards Under the Health Insurance Portability and Accountability Act of 1996 (HIPAA)

Patricia L. Watts, J.D., Denise M. Hynes, Ph.D., and April Kopp

Introduction

In this issue of *VIReC Insights*, we provide a brief background of HIPAA, describe the Privacy Rule, and review the VHA research implications while highlighting and defining key terms and relevance to the research process. We conclude with specific references and Web citations for additional information.

Background

The purpose of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 (Public Law 104-191) is to improve portability and continuity of health insurance coverage in the group and individual markets. A key element in accomplishing this objective is the simplification of the administration of health insurance. The

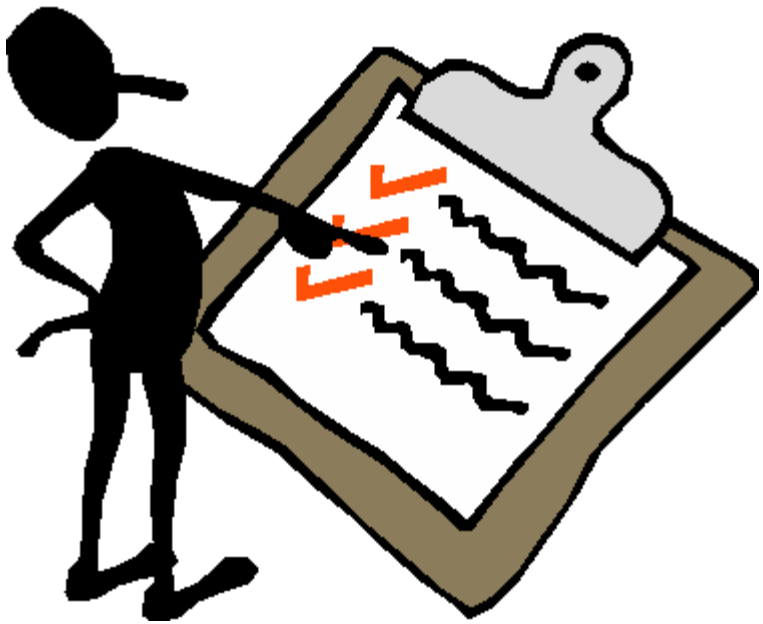
Additional Information

- On HIPAA Administrative Simplification and DHHS Proposed Rules: http://aspe.hhs.gov/admnsimp and http://www.cms.hhs.gov/hipaa/hipaa2/default.asp
- VHA HIPAA Effort: HIPAA PMO http://vaww1.va.gov/cbo/hipaa.html
- Or contact HIPAA PMO at 202-254-0385

HIPAA Data Security Standards

- HIPAA security standards apply to all health information pertaining to an individual that is electronically maintained or electronically transmitted
- Must protect against unauthorized access and misuse of electronic health information

What Is Required?



- Protect data
  - at home facility
  - in transit
  - at recipient facility
- Using safeguards
  - administrative
  - physical
  - technical

Home Facility: Administrative Safeguards

- Establish a Data Use Policy:
  - Who's in charge of data security?
  - How are users granted access?
  - How is access limited to authorized users?
  - How is authorization terminated?
  - What do you do if security fails?

Home Facility: Physical Safeguards

- Where data are stored (room/server)
- When data are displayed on monitor
- When data are on portable media (laptops/CDs)

Home Facility: Technical Safeguards

- Control access
  - passwords
  - encryption
  - authentication

Security of Data In Transit

- Administrative
  - Virtual Private Network
  - Privacy warning labels
- Physical
  - Courier with ground tracking
- Technical
  - Zip/Encrypt
  - Password protection
  - Authentication software

Security of Data At Recipient Facility

- Assurances regarding:
  - Administrative Safeguards
  - Physical Safeguards
  - Technical Safeguards
- Data Use Agreement (DUA)
  - Not currently required for research if within the VHA but may still be encountered from some offices
  - If not within VHA…

Get To Know Local Data Security Resources

- Privacy Officer
- Information Security Officer
- Information Resources Management (IRM) Office

National Resources

- Office of Cyber Security
  - http://infosec.va.gov/main/index.asp
- HIPAA Administrative Simplification and DHHS Proposed Rules
  - http://aspe.hhs.gov/admnsimp and http://www.cms.hhs.gov/hipaa/hipaa2/default.asp
- OI HIPAA Effort: http://vaww.va.gov/hipaa
- Stephania Putt, OI HIPAA Coordinator
  - (727) 320-1839

Questions ?

7th Annual Epidemiology, Biostatistics and Clinical Research Methods Summer Session

June 20-24, 2005

Inpatient and Outpatient Data Sets

Session Objectives

- Become aware of utility of VA data sources for cancer epidemiology and health services research
- Become aware of the scope and breadth of VA data
- Know about *G*ood *D*ata *P*ractices
- **Understand how to use the VA Inpatient & Outpatient datasets for research**
- Describe an example of determining cancer prevalence using VA data
- Know where to go for help when using VA data

VA Inpatient and Outpatient Medical SAS Datasets

- **National** VHA health care delivery data
- **Administrative/workload** purposes
- **SAS** datasets housed on mainframe computer at **Austin Automation Center (AAC)**
- Essential for examining VA health services use

Inpatient Datasets

Inpatient Datasets

- Medical SAS Inpatient Datasets
  - Often referred to as the **PTF** or Patient Treatment Files
  - Note: PTF in VISTA is not the same
- Record created at Discharge
- Admission may be previous year
- Types of stays:
  - Acute Care
  - Extended Care
  - Observation Care
  - Non-VA Care

Inpatient Data Flow

Data Flow from the VHA Medical Centers to the Austin Automation Center (AAC) to the Medical SAS Inpatient Datasets

*VHA Medical Centers (local **VISTA** system)*

PTF

Acute Care: Main, Procedure, Bed Section, Surgery Datasets

Extended Care: Main, Procedure, Bed Section, Surgery Datasets

Observation Care: Main, Procedure, Bed Section Datasets

Non-VA Care: Main, Procedure, Bed Section, Surgery Datasets

Medical SAS Inpatient Datasets

Inpatient SAS Datasets: For Each Type of Care

File	Reference	Dates
Main	PM	1970 - present
Bedsection	PB	1984 – present
Procedure	PP	1988 – present
Surgery	PS	1984 - present

Data Elements In All Inpatient Datasets

- Patient identifier (SCRSSN)
- Facility & VISN identifiers of where care was provided
- Admission & Discharge Date & Time
- Discharge Type (e.g., Regular, Death-Autopsy, Non-bed Care)
- Principal Diagnosis (DXPRIME) for admission

Special Topic: SCRSSN

- Scrambled Social Security Number
- Same algorithm used for all Inpatient and Outpatient Care SAS Datasets
- Link patient over time and across datasets
- Real SSN available with authorization

Special Topic: Diagnosis

- DXPRIME = Principal Diagnosis
  - Condition determined to be chiefly responsible for the admission
- DXLSF = Primary Diagnosis
  - Diagnosis responsible for the major part of the full stay

Inpatient Main Dataset

Dataset	Reference	Dates
Main	**PM**	**1970 - present**
Bedsection	PB	1984 – present
Procedure	PP	1988 – present
Surgery	PS	1984 - present

Inpatient Main Dataset Overview

- One record per discharge
- Diagnostic codes in ICD-9
  - Principal Diagnosis: DXPRIME
  - Secondary Diagnoses: DXF2 – DXF10
  - Additional diagnoses in Bedsection Dataset
- Patient demographics
  - Age, sex, race, DOB, DOD, state and ZIP code of residence
- Care characteristics
  - LOS, discharge disposition, DRG
- Service experiences
  - AOR, Combat, POS, POW, RAD, SCI, SCPER

Special Topic: Race

- New standard implemented in May 2003
- Compliant with 1997 OMB Directive
- New standard:
  - Self-identification is preferred
  - Multiple race reporting allowed
  - Race and ethnicity separately collected
- Since FY2003, Main Datasets include:
  - RACE: old format (six categories)
  - RACE1 – RACE6: race with collection method
  - ETHNIC: ethnicity with collection method

Inpatient Bedsection Dataset

Dataset	Reference	Dates
Main	PM	1970 - present
Bedsection	**PB**	**1984 – present**
Procedure	PP	1988 – present
Surgery	PS	1984 - present

Bedsection Dataset Overview

- Record = Bedsection stay
- Bedsection = treating specialty of physician, not physical location
- Maximum of 25 bedsection records per inpatient stay

Bedsections – Treating Specialties

Cardiology	**Cardiac Step-Down**	**Medical ICU**
Surgery (Gen)	**Surgical ICU**	**Surgical Obs**
Spinal Cord Injury	**SCI Observation**	**Orthopedic**

See full listing in *VIReC Research User Guide: FY2002 VHA Medical SAS® Inpatient Datasets*

Inpatient Bedsection Dataset Selected Data Elements

- Date & time of transfer into & out of Bedsection
  - Use BSINDAY & BSOUTDAY to compute acute care LOS
- Physical Location Code (PLBED)
- Bedsection Diagnosis (max. 5) & Bedsection DRG

Inpatient Bedsection Dataset Selected Data Elements

- Service-Connected Treatment
  - Is the condition being addressed in the bedsection a service-connected one?
  - Not the same as a veteran's service-connected eligibility for mandatory care (e.g., Agent Orange exposure)

Inpatient Procedure Dataset

Dataset	Reference	Dates
Main	PM	1970 - present
Bedsection	PB	1984 – present
Procedure	**PP**	**1988 – present**
Surgery	PS	1984 - present

Inpatient Procedure Dataset Selected Data Elements

- Procedure, coded in ICD-9-CM (vs. CPT for outpatient procedures)
- Dialysis type & number of dialysis treatments
- Physician's specialty (bedsection)

Inpatient Surgery Dataset

Dataset	Reference	Dates
Main	PM	1970 - present
Bedsection	PB	1984 – present
Procedure	PP	1988 – present
Surgery	**PS**	**1984 - present**

Special Topic: Procedures vs. Surgeries

- Surgery = Procedure performed in main or specialized operating room
- "Procedure" in Facility A may = "Surgery" in Facility B
  - Depends on where performed
- Look at both datasets

Outpatient Datasets

Outpatient Datasets

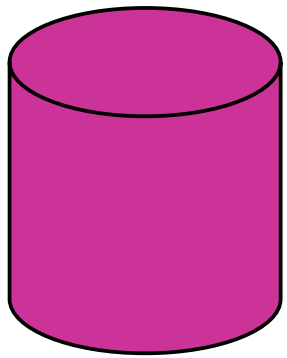
- Referred to as OPC or NPCD (National Patient Care Database)

- 4 datasets:
  - Visit
  - Procedure
  - Diagnosis
  - Event

Outpatient Data Flow

Data Flow from the VHA Medical Centers to the Austin Automation Center (AAC) to the Medical SAS Outpatient Datasets

*VHA Medical Centers (local **VISTA** system)*



Tables at AAC (relational database)

Visit Dataset – Clinic stop codes and patient demographics for all encounters

Event Dataset – Procedure codes, diagnosis codes and patient demographics for one encounter at one clinic

Medical SAS Outpatient Datasets

Outpatient Datasets

File	Reference	Dates
Visit	**SF**	**1980 - present**
Event	**SE**	**1998 – present**
Diagnosis	SG	1997 – 2001
Procedure	SC	1990 - 2001

Data Elements Common In Outpatient Datasets

- Patient identifier (SCRSSN)
- Patient demographics (Age, date of birth, race, marital status)
- Patient Zip Code, County, & State of Residence
- Date of encounter

Common Data Elements

- Means Test Indicator (MEANS)
- Patient eligibility code (ELIG)
  - ELIG < 11 identifies a veteran
- Agent Orange exposure claimed (AOIND/AGOLOC)
- Radiation exposure claimed (RAD)
  - Visit: 4 categories
  - Event: exposed vs. not exposed

Outpatient Visit Dataset

- Record = One day's encounters for a patient at a clinic

Outpatient Visit Dataset Selected Data Elements

- One record per visit
- Up to 15 clinic stops per visit
- No diagnosis or procedure information
- Race:
  - RACE
  - RACE1 – RACE7 from FY2004
- Insurance and religious preference only in Visit file

Outpatient Event Dataset

- Record = Ambulatory encounter
  - Coded as Primary Clinic Stop, which is officially referred to as DSS Identifier
- No limit on number of encounters per day

Outpatient Event Dataset Selected Data Elements

- Record = Ambulatory encounter or clinic stop
- No limit on number of clinic stops per day
- Up to 10 diagnoses in ICD-9 codes (DXLSF, DXF2 – DXF10) per record
- Until FY2003: Up to 15 procedures in CPT4 codes (CPT1 – CPT15), no repeats allowed
- Since FY2004: Up to 20 CPT4 codes with repetition allowed
- Encounter ID to link the Event dataset with HERC Outpatient Average Cost Dataset since FY2003
- Appointment type only in Event Dataset

Strengths & Limitations of Medical SAS Datasets

Strengths

- Centralized data source
- Large groups of patients
- Given good coding, reflective of general clinical status
- Unique identifier (SCRSSN) allows linking records across files/years

Limitations

- Not all care dimensions
- Retrospective discharge abstracts
- Incentives to coding
- Limitations of ICD-9-CM coding

Some Frequently Asked Questions (FAQs)

- Question:
- How reliable is the race variable in the Medical SAS Datasets?

- Answer:
- Reliable (~92% agreement with Medicare race in identifying African-American race).
- FY2003 data: 90% missing. Consider obtaining race from other sources

FAQs

- Question:
- How reliable is ICD-9 coding in Medical SAS Datasets?

- Answer:
- Varies: 54% (stroke) – 98% (cardiac)
- Overcoding: hypertension (31%) diabetes (19%)

http://www.measurementexperts.org/learn/practice/ab_icd-9_pf.asp

FAQs

- Question:
- Why do I need to use DXPRIME instead of DXLSF?

- Answer:
- DXPRIME is the condition determined to be chiefly responsible for the admission & differs from the DXLSF

FAQs

- Question:
- How do I compute acute length of stay (LS)?

- Answer:
- LS in Main may include extended care stay
- Use Bedsection data (BS in and out day) to compute acute LOS
- HERC inpatient Average Cost Datasets documentation has details (pp. 29 - 32)

Data Quality Information

- Quality assessments performed by the Office of Inspector General, the Medical Care Cost Recovery program, and special workgroups

- Data Quality, Information Assurance, Office of Information (http://vaww.vhaco.va.gov/dataquality/default.htm).

Data Quality Information (cont'd)

- VHA Coding Council (http://vaww1.va.gov/health/him/VHACC/VA_HIM_P/coding_council1.htm) - VHA Coding Handbook
- HSRData e-mail listserv

Session Objectives

- Become aware of utility of VA data sources for cancer epidemiology and health services research
- Become aware of the scope and breadth of VA data
- Know about *G*ood *D*ata *P*ractices
- Understand how to use the VA Inpatient & Outpatient datasets for research
- **Describe an example of determining cancer prevalence using VA data**
- Know where to go for help when using VA data

Prostate Cancer Prevalence Example

- Prevalence of prostate cancer in FY2003
  - VHA Users?
  - Veterans
- Demographics
  - Age
  - Race
  - Region
  - Insurance

Prevalence of Prostate Cancer: Linking Datasets

- Can be linked forward or backward in time
  - Utilization
  - Treatment
  - Comorbidities
  - Complications
- Other databases
  - Pharmacy
  - DSS LAR (e.g., for PSA Test Results)
  - Medicare Data

Prevalence of Prostate Cancer: Datasets & Variables

- FY2003 Inpatient and Outpatient Datasets
  - Inpatient datasets
    - Acute care only
  - Outpatient datasets
- Diagnostic codes only
- Demographic information
  - Age and race, both in- and outpatient datasets
  - Insurance, only in Outpatient Visit dataset
  - Region, constructed using STA6A and STA5A and PSSG files

Prevalence of Prostate Cancer

(All Males 40 and Older, FY2003 – continued)

	Unweighted Frequency (# cases)		Base Population	Rate per 100,000	
	Primary Diagnosis	All Diagnoses		Primary Diagnosis	All Diagnoses
Total	113,948	191,513	4,474,394	2,546.67	4,280.20
Insurance Status					
No ins/self-pay	45,445	69,903	2,275,112	1,997.48	3,072.51
Medicare	53,566	96,495	1,489,789	3,595.54	6,477.09
Medicaid	192	234	9,170	2,093.78	2,551.80
Private Ins/HMO	14,177	24,053	664,958	2,132.01	3,617.22
Other Insurance	549	775	33,763	1,626.04	2,295.41
Unknown	19	53	1,602	1,186.02	3,308.36
Region					
Eastern	19,802	37,175	760,864	2,602.57	4,885.89
Central	22,637	40,704	997,702	2,268.91	4,079.78
Southern	47,199	77,210	1,820,440	2,592.72	4,241.28
Western	24,310	36,424	895,388	2,715.02	4,067.96

Prevalence of Prostate Cancer

(All Males 40 and Older, FY2003)

	Unweighted Frequency (# cases)		Base Population	Rate per 100,000	
	Primary Diagnosis	All Diagnoses		Primary Diagnosis	All Diagnoses
Total	113,948	191,513	4,474,394	2,546.67	4,280.20
Age					
40 - 44	142	191	203,382	69.82	93.91
45 - 54	4,511	5,606	855,506	527.29	655.28
55 - 64	18,074	24,060	999,578	1,808.16	2,407.02
65 - 74	41,018	68,947	1,202,705	3,410.48	5,732.66
75 - 84	44,964	83,348	1,085,096	4,143.78	7,681.16
85 +	5,239	9,361	128,127	4,088.91	7,306.03
Race-Ethnicity					
White	80,681	146,862	3,014,012	2,676.86	4,872.64
Black	21,975	28,183	446,750	4,918.86	6,308.45
Hispanic	2,282	3,105	92,705	2,461.57	3,349.33
Other	1,189	1,739	46,985	2,530.59	3,701.18
Unknown	7,821	11,624	873,942	894.91	1,330.07

Session Objectives

- Become aware of utility of VA data sources for cancer epidemiology and health services research
- Become aware of the scope and breadth of VA data
- Know about *G*ood *D*ata *P*ractices
- Understand how to use the VA Inpatient & Outpatient datasets for research
- Describe an example of determining cancer prevalence using VA data
- **Know where to go for help when using VA data**

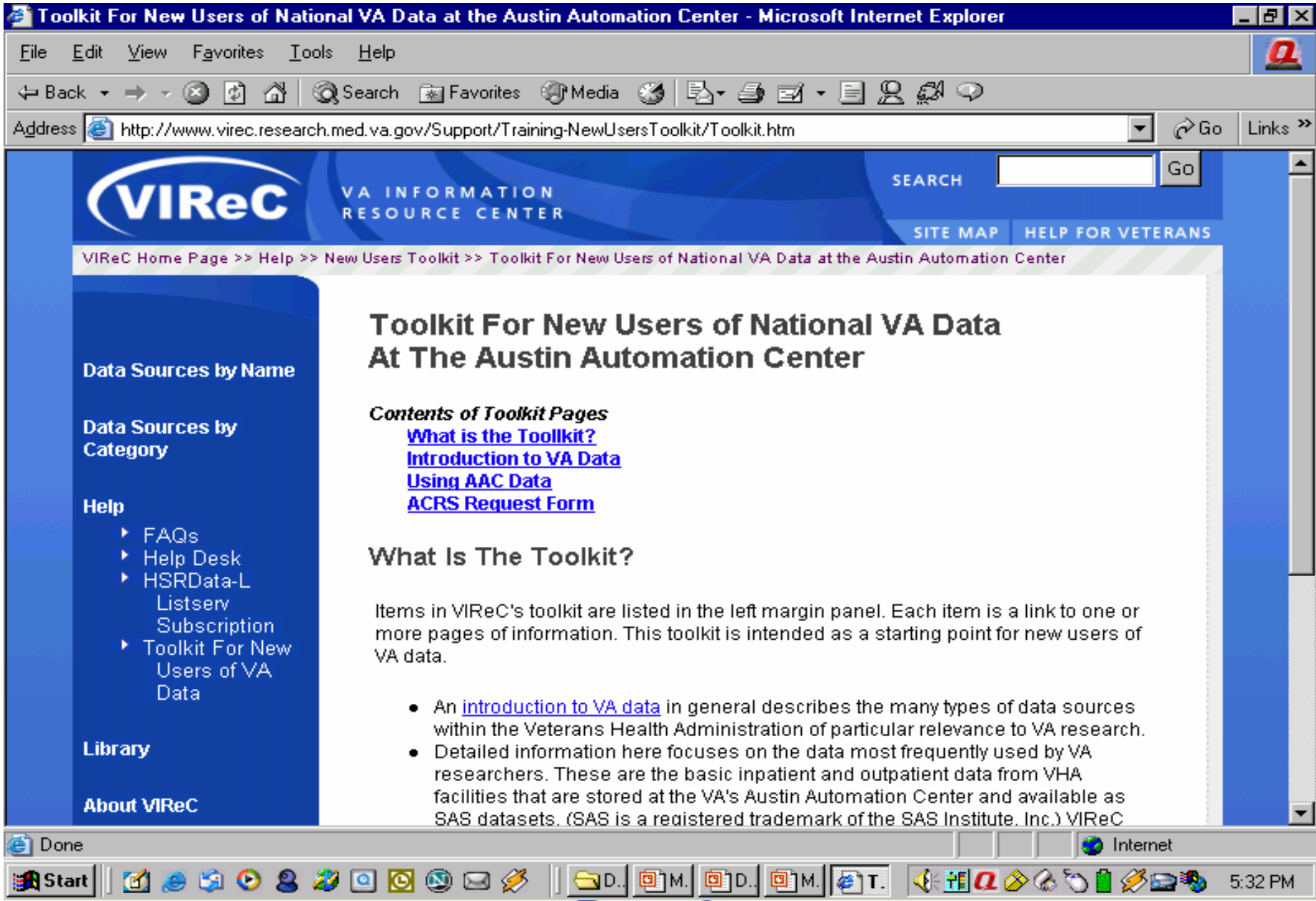
Obtaining Help Using VA Data

- AAC Help Desk
  - (512) 326-6780
  - Questions about access to data at AAC
  - Help with file names, SAS programming

Other Sources of Help

- HSRData Listserv
  - Join at VIReC Web site
  - Discussion among > 200 data stewards, managers, and users
  - Past messages in archive

- VIReC Toolkit for New Data users
  - http://www.virec.research.med.va.gov/Support/Training-NewUsersToolkit/Toolkit.htm


Toolkit For New Users of National VA Data at the Austin Automation Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.virec.research.med.va.gov/Support/Training-NewUsersToolkit/Toolkit.htm
Go Links
VIReC
VA INFORMATION RESOURCE CENTER
SEARCH
Go
SITE MAP
HELP FOR VETERANS
VIReC Home Page >> Help >> New Users Toolkit >> Toolkit For New Users of National VA Data at the Austin Automation Center
Data Sources by Name
Data Sources by Category
Help
FAQs
Help Desk
HSRData-L Listserv Subscription
Toolkit For New Users of VA Data
Library
About VIReC
Toolkit For New Users of National VA Data At The Austin Automation Center
Contents of Toolkit Pages
What is the Toolkit?
Introduction to VA Data
Using AAC Data
ACRS Request Form
What Is The Toolkit?
Items in VIReC's toolkit are listed in the left margin panel. Each item is a link to one or more pages of information. This toolkit is intended as a starting point for new users of VA data.
An introduction to VA data in general describes the many types of data sources within the Veterans Health Administration of particular relevance to VA research.
Detailed information here focuses on the data most frequently used by VA researchers. These are the basic inpatient and outpatient data from VHA facilities that are stored at the VA's Austin Automation Center and available as SAS datasets. (SAS is a registered trademark of the SAS Institute, Inc.) VIReC
Done
Internet
Start
5:32 PM

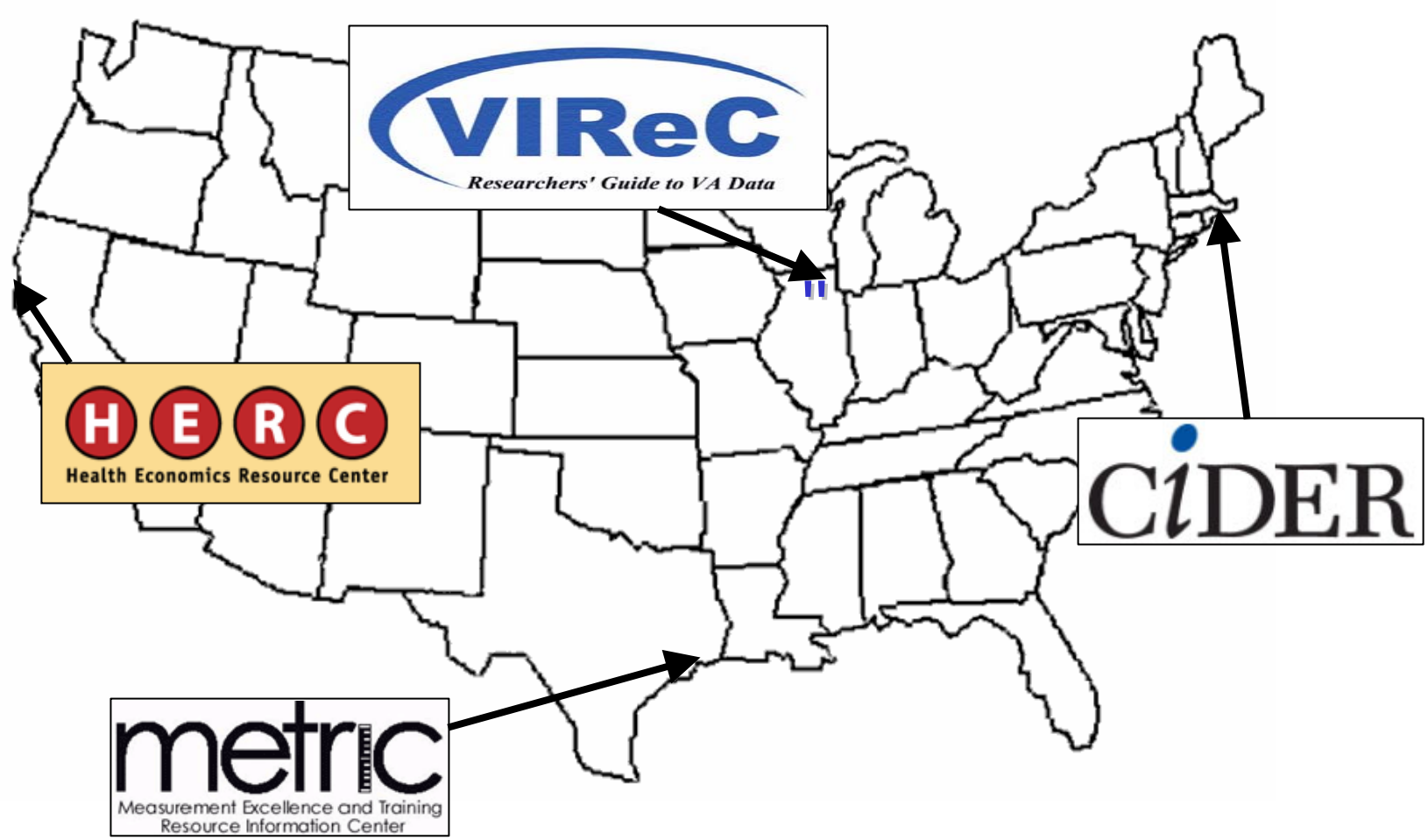
Other Sources for Help

- VHA MetaData Registry
  - Search for data element
  - Database
  - Valid codes
  - Database contact
  - In development and expanding
- http://vaww.metadata.med.va.gov

VIReC Help

- VIReC Help Desk
  - VIReC staff will answer your question and/or direct you to available resources on topics
  - http://www.virec.research.med.va.gov
  - VIReC@va.gov
  - (708) 202-2413

VA HSR&D Service Resource Centers



Questions ?